Environment-wide association study (EWAS) to identify factors associated with hematocrit: evidence from the Guangzhou Biobank Cohort Study

Y, Zhong,; Jiang, C. Q.; Cheng, Kar; Zhang, W. S.; Jin, Y L; Lam, TH; Woo, J; Leung, GM; Schooling, CM

DOI:
10.1016/j.annepidem.2016.07.005

Document Version
Peer reviewed version

Citation for published version (Harvard):
Y, Z, Jiang, CQ, Cheng, K, Zhang, WS, Jin, YL, Lam, TH, Woo, J, Leung, GM & Schooling, CM 2016, 'Environment-wide association study (EWAS) to identify factors associated with hematocrit: evidence from the Guangzhou Biobank Cohort Study', *Annals of Epidemiology*, vol. 26, no. 9, pp. 638-642. https://doi.org/10.1016/j.annepidem.2016.07.005

Environment-wide association study (EWAS) to identify factors associated with hematocrit: evidence from the Guangzhou Biobank Cohort Study

Zhong Y [1], Jiang CQ[2], Cheng KK[3], Zhang WS[2], Jin YL[2], Lam TH[1], Woo J,[4] Leung GM[1], Schooling CM[1,5]

[1] School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong

[2]Guangzhou No. 12 Hospital, Guangzhou, Guangdong, China

[3]Department of Public Health and Epidemiology, University of Birmingham, Birmingham, UK

[4]Department of Medicine & Therapeutics, The Chinese University of Hong Kong

[5]CUNY School of Public Health and Hunter College, New York, USA

Corresponding author: Lam TH

Address: 5/F William MW Mong Block, 21 Sassoon Road, Hong Kong.

Tel: (852)39179287 Fax: (852)28559528 Email: hrmrlth@hku.hk

Running head: Environmental factors associated with hematocrit

Word count (abstract): 194

Word count (main text): 2,739

Keywords: cardiovascular disease; Chinese; coagulability; environment-wide association study; hematocrit;

Purpose: In randomized controlled trials reducing high hematocrit (Hct) in patients with polycythemia vera protects against CVD events, whilst increasing Hct in anemia patients causes CVD events. Hct is influenced by environmental and life style factors. Given the limited knowledge concerning the drivers of Hct, we took an approach to identifying drivers of Hct.

Methods: We used an environment-wide association study (EWAS) to identify environmental and life style factors associated with Hct in 20443 older Chinese adults (mean age=62.7 years) from the Guangzhou Biobank Cohort Study. We evaluated the role of 25 nutrients, 40 environmental contaminants, 2 metals (only available for 10405 participants) and 6 life-style factors in relation to Hct, adjusted for sex, age, recruitment phase and social-economic position.

Results: In a mutually adjusted model vitamin A, serum calcium, serum magnesium and alcohol use were associated with higher Hct while physical activity was associated with lower Hct.

Conclusion: Despite the difficulty of ascertaining causality, finding both expected (vitamin A and physical inactivity) and novel factors (serum calcium, serum magnesium and alcohol use) strongly associated with Hct illustrates the utility of EWAS to generate hypotheses regarding the potential contribution of modifiable exposures to CVD.

Abbreviations:

CI – confidence interval

CVD – cardiovascular disease

EWAS – Environment-wide association study

FDR – false discovery rate

GBCS – Guangzhou Biobank Cohort Study

GHHARE – Guangzhou Health and Happiness Association for the Respectable Elders

GWAS – Genome-wide association studies

G6PD – Glucose-6-phosphate dehydrogenase

Hct – hematocrit

HEPA – health enhancing physical activity

Hgb – hemoglobin

PTH – parathyroid hormone

RCT – randomized controlled trial

SD – standard deviation

With improving living standards and socioeconomic development, cardiovascular disease (CVD) has replaced infectious diseases as the leading cause of death both in developed and developing countries (1). Substantial progress has been made in teasing out the underlying causes of CVD from numerous risk factors. However, recently unexpected findings from randomized controlled trials (RCT) have revealed that some CVD risk factors are likely non-causal, such as HDL-cholesterol (2, 3), C-reactive protein (4) and fasting glucose (5). As such, all three factors of Virchow's triad described over 100 years ago might be relevant to CVD prevention, i.e., flow (hemodynamics), vessel (endothelial injury) and blood (hypercoagulability including viscosity). In the past century, risk factors for CVD from hemodynamic

changes or endothelial injury, such as hypertension or higher LDL, have been extensively investigated. However, the impact of viscosity has been less comprehensively assessed, despite anti-coagulants, such as warfarin, being a mainstay of CVD treatment (6).

Viscosity has been thought to be a risk factor in cardiovascular disease (CVD) since the 1960s (7) and Hct is one of the major determinants of viscosity. (8) Several longitudinal studies have found an association of high hematocrit (Hct) with ischemic heart disease (IHD). (9, 10) Such associations may not always be evident because they may be biased towards the null by reverse causality as Hct falls with ill-health. (11) The same association of higher Hct with higher risk of myocardial infarction has also been seen among young Swedish men. (12) Genome-wide association studies (GWAS) are beginning to identify genes connected with viscosity and coagulability as related to CVD. (13) Notably, carriers of one thalassemia allele are more prone to anemia, but less vulnerable to CVD. (14) Similarly, people with Glucose-6-phosphate dehydrogenase (G6PD) deficiency are at risk of hemolytic anemia in states of oxidative stress, but have a lower risk of ischemic heart disease and cardiovascular associated death. (15-17) Experimental evidence also suggests that Hct and /or hemoglobin (Hgb) may play a causal role in CVD. In randomized controlled trials (RCTs) treatment of anemia with a higher Hgb target increased hospitalization and mortality rates from CVD events in patients with chronic kidney disease without improving quality of life. (18) Similarly, in an RCT, polycythemia vera patients treated with a lower Hct target had a lower rate of cardiovascular death and major thrombosis than those with a higher Hct target, suggesting reducing Hct could protect against CVD events. (19) In addition, several longitudinal studies suggest that Hct in the normal range is positively associated with CVD. (9, 10, 12) Taken together this evidence suggests that Hct might have a causal effect on CVD across the normal range, although an experimental or Mendelian randomization study would be necessary to confirm or refute that hypothesis. Moreover, the observed association of Hct with CVD suggests that if Hct does not play a causal role in CVD then a confounder which causes both Hct and CVD should exist, which may be identified from a systematic search of factors associated with Hct.

Nevertheless, research on the drivers of Hct is scarce, although these could be

potential targets for CVD prevention and/or treatment. Hct is partly determined by genetic factors, but environmental, dietary, behavior and life-course factors influences account for 35%-60% of the total variance. (12) Classic epidemiologic studies typically test a limited number of factors at a time, focusing on corroborating or refuting the current paradigm, somewhat akin to candidate gene studies. However, genome-wide analysis study (GWAS) has revealed unexpected associations with disease and genetic variants with unknown function as associated with cardiovascular disease, suggesting the existence of many overlooked pathways. An agnostic approach may yield new insights. To identify potential environmental drivers of Hct, we used a systematic approach to test the association of multiple environmental, dietary and lifestyle factors with Hct, i.e., an environment-wide association study (EWAS) analogous to a GWAS, (20) in a large well-characterized sample of southern Chinese.

METHODS

Ethics statement

All participants had given written informed consent before participating. The study has ethics approval from the Guangzhou Medical Ethics Committee of the Chinese Medical Association, Guangzhou, China.

Sources of data

The Guangzhou Biobank Cohort Study (GBCS) is a collaboration among the Guangzhou No. 12 Hospital and the Universities of Hong Kong and Birmingham, UK. (21) Participants were recruited from the Guangzhou Health and Happiness Association for the Respectable Elders (GHHARE), a community social and welfare association unofficially aligned with the municipal government, where membership is open to anyone aged 50 years or older for a monthly, nominal fee of 4 Yuan (50 US cents). Recruitment for phase 1 took place from September 2003 to November 2004, phase 2 from April 2005 to May 2006 and phase 3 from September 2006 to January 2008. Follow-up of the participants started in 2008. Approximately 7% of permanent Guangzhou residents aged 50 years or more are members of GHHARE, of whom 33% enrolled for phases 1, 2 or 3 recruitment and were included if they were capable of consenting, ambulatory and not receiving treatment modalities that, if omitted,

might result in immediate life-threatening risk, such as chemotherapy or radiotherapy for cancer or dialysis for renal failure. Participants included in this analysis underwent a half-day detailed medical interview, which included a physical examination and a questionnaire that asked about disease history, health-related habits and demographic characteristics (21). We used only phase I and II data because a different dietary instrument was used in phase III. (22)

The questionnaire asked about exposure to occupational health hazards (including dust/gas or chemical fumes/physical exposures) during their longest held job, voluntary job or at other times, and the duration (how many years) and level (mild/moderate/severe) of exposure. Dietary intake was assessed using a validated 7-day food frequency questionnaire. (23) The food frequency questionnaire consisted of items in the following seven categories: Bread/pasta/rice (22 items); vegetables (66 items); fruits (30 items); meat (42 items)/fish (37 items)/eggs (8 items); beverages (36 items); dimsum/snacks (54 items); soups (8 items); and oil/salt/sauces. Items chosen were those most frequently consumed, based on previous surveys in southern China. Each participant reported for each food item, the portion size and the number of times of consumption each week. Portion size was explained using a catalogue of pictures of individual food portions. (24) Fasting blood samples were drawn using a vacuum tube and tested in the hospital laboratory. A complete blood count was performed with a SYSMAX K-21 clinical chemical analyzer including hematocrit.

Exposure assessment

We obtained 41 environmental hazards from the questionnaire, defined as exposure level multiplied by exposure time. We obtained 25 nutrients, including dietary intake of total calories, macronutrients, vitamins and trace elements calculated from the food frequency questionnaire and dietary composition tables based on McCance and Widdowson (22) and the Chinese Medical Sciences Institute (25). We also obtained 6 key lifestyle attributes, physical activity, alcohol use, smoking status, sleep hours and meat eating habits or vegetarianism. Physical activity was categorized as: inactive, minimally active and health enhancing physical activity (HEPA). Alcohol use was categorized as never, less than once per week, 1-4 times per week, more than 5

times per week, ex-drinkers or unknown. Smoking status was defined as never-smoker, ex-smoker or current smoker. Meat eating habits were defined as: "never", "about once a year", "about once a month", "about once a week" or "almost daily". Two metals, calcium and magnesium, were measured in whole blood in phase 1 only (hence, data only available for 10405 participants).

Outcome measures

The primary outcome was Hct, considered as a continuous variable in the units in which it was measured.

Statistical analysis

All continuous variables were transformed into z-scores to facilitate comparability between variables. Some lifestyle factors were on an ordinal scale (exercise 3 levels, drinking 6 levels, smoking 3 levels, meat eating habits 5 levels). First, a principal component factor analysis was used to investigate the interrelationship between the 74 environmental and life lifestyle variables. Second, linear regression was used to assess the adjusted associations of each of the environmental and lifestyle variables with Hct adjusted for sex, age, recruitment phase and social-economic position (income, longest held occupation and education). One problem with hypothesis-free, data-driven research, such as EWAS, is the validation of positive findings. We used internal cross-validation with 5 iterations and only 80% of the sample in each run, repeated five times with a different quintile excluded each time. Variables with a false discovery (FDR) <5% in all 5 runs were retained. The FDR is the expected proportion of false positives. An FDR <5% was used based on the Bonferroni method. (26) Finally, variables with an FDR<5% in all 5 runs were considered together, excluding any variables closely related to each other in the initial factor analysis, with the confounders in a multiple linear regression model with Hct as the dependent variable.

Results

Of the 20,415 participants recruited and examined, in phase I and II of the Guangzhou Biobank Cohort Study, 99.4% (20,298) had complete data on hematocrit

and potential confounders with dummy categories used for the 13.3% with missing job type and 4.8% with missing personal income. The numbers of exposures varied mostly from 19024 to 20415 except for phase I data in which serum calcium and magnesium were available. Means and medians for the nutrients, hazardous exposures and lifestyle factors are shown in Appendix Table A.

Associations for the 74 environmental contaminants and lifestyle factors with Hct adjusted for sex, age, recruitment phase and social-economic position in the total sample are shown in Figure 1 with p-values on the log 10 scale. Fourteen variables were identified with an FDR <5% in the total sample. However, only 6 variables (vitamin A intake, vitamin C intake, serum calcium, physical activity, serum magnesium and alcohol use) showed an FDR<5% in all 5 runs of the cross-validation procedure. Together in a multivariable linear regression model, vitamin A intake, serum calcium, serum magnesium and alcohol use were associated with higher Hct while physical activity was associated with lower Hct (Table 1). However, vitamin C intake was not associated with Hct. Appendix Figure A shows the specific coefficients and 95% confidence intervals for all 74 factors adjusted for sex, age, recruitment phase and social-economic position.

Table I **Adjusted[a] associations of the six variables with FDR<5% in all 5 runs in the cross-validation procedure with Hct in the Guangzhou Biobank Cohort Study, Guangzhou, China, 2003-2006**

Variable	Z-score results		Original results	
	Coefficient	95% CI	Coefficient	95% CI
Vitamin A (ug/day)	0.36	(0.28, 0.44)	0.0007	(0.0006, 0.0008)
Serum calcium (mmol/L)	0.41	(0.34, 0.48)	2.27	(1.86, 2.68)
Physical activity	-0.25	(-0.33, -0.17)	-0.25	(-0.33, -0.17)
Vitamin C (mg/day)	-0.04	(-0.13, 0.04)	-0.003	(-0.01, 0.004)
Serum magnesium (mmol/L)	0.19	(0.13, 0.26)	0.89	(0.54, 1.24)
Alcohol use	0.20	(0.13, 0.26)	0.20	(0.13, 0.26)

[a] Adjusted for age, sex, recruitment phase and social-economic position (income,longest held occupation and education)

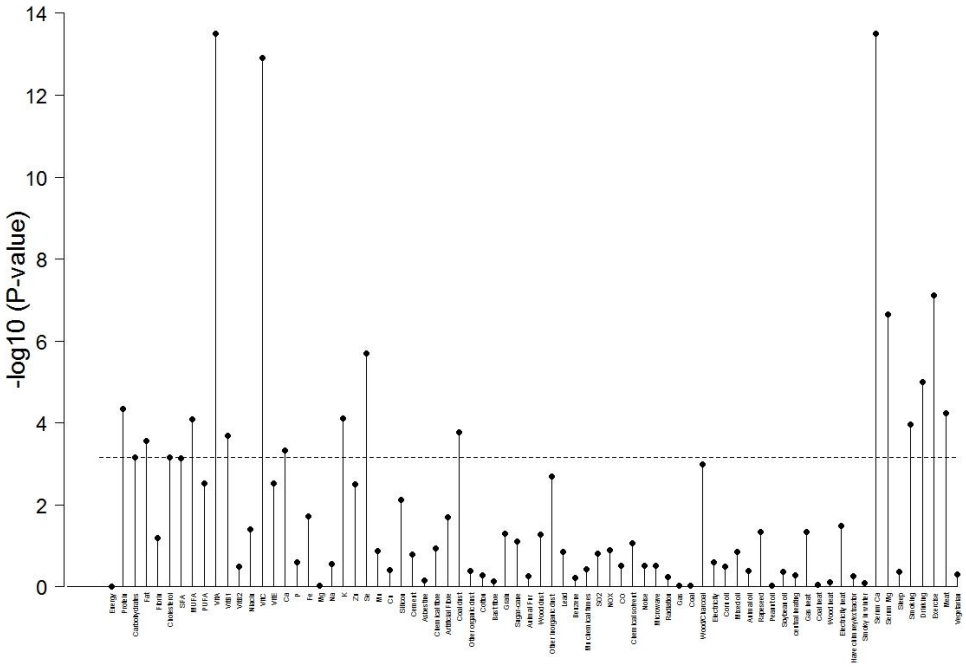


Fig.1. “Manhattan” type plot showing the p-values (expressed at the –log 10 scale) in the total sample for 74 nutrients, environment contaminants or lifestyle factors with Hct values in the Guangzhou Biobank Cohort Study , Guangzhou, China, 2003-2006. The broken line denotes the cutoff for a false discovery rate (FDR) <5%. For abbreviations of variables, see Appendix Table A

Discussion

The role of environment and lifestyle factors in Hct has not previously been examined systematically and comprehensively. This study used an EWAS approach to determine potential drivers of Hct. We found both known drivers of Hct or hemoglobin, i.e., vitamin A and physical inactivity, and unknown potential drivers of Hct, i.e., serum calcium, serum magnesium and alcohol use. Interventions have shown that vitamin A raises Hct (27-30) and physical inactivity reduces Hct (31-34).

Vitamin A may reduce anemia via a number of mechanisms. Vitamin A may increase resistance to some specific infections, such as measles or diarrhea. (35) Vitamin A may directly modulate erythropoiesis, such as retinoids regulating programmed cell death in erythroid progenitor cells (36). Vitamin A may act via other factors such as testosterone, which is a routine therapy for aplastic anemia. (37, 38) Bringing together these disparate findings provides a potential explanation for high levels of vitamin A increasing mortality, (39, 40) because vitamin A increases Hct, which in turn increases the risk of cardiovascular disease. (18, 19)

Intervention trials (31-34) also suggest that moderate physical activity lowers Hct because moderate exercise intensity is followed by activation of blood fibrinolysis without concomitant hypercoagulability, whereas strenuous exercise activates both the fibrinolysis and hypercoagulability systems simultaneously. (41)

Serum calcium was positively associated with Hct. Serum calcium could affect Hct directly or indirectly, via raising some other factor that affects Hct. Alternativeky, a common factor, such as parathyroid hormone (PTH) or chronic kidney disease could underlie the association, by affecting calcium and Hct. PTH regulates serum calcium and in animal experiments administration of PTH increases hematopoietic stem cells (42, 43), whereas chronic kidney disease reduces serum calcium and erythropoietin. PTH was not measured in GBCS, so we cannot assess this possibility. Adjusting for a

marker of renal function (serum creatinine) did not change the association, indicating that renal function may not be a confounding underlying the association.

Magnesium was once hypothesized to protect against some cardiovascular conditions, (44) however that hypothesis has been refuted. (45) Whether any beneficial effects of magnesium have been offset by raising Hct does not seem to have been considered, and evidence is limited. Alcohol use was associated with higher Hct perhaps because the oxides or toxins from alcohol cause hypoxia which increases Hct. (46, 47)

Smoking is known to raise Hct. However, smoking did not cross the FDR threshold in one internal cross-validation, although it had a low p-value. Smoking is uncommon in women in Southern China (48), and men smoke less than men in the west with many occasional smokers (49) which may explain why smoking was not very strongly associated with Hct in this study. Moreover, the EWAS method is not intended to find all possible correlations, but just to identify the stronger associations that might be most relevant in a specific population and worthy of further study.

Several limitations bear mention. First, Hct could be influenced by recent nutrition or frequency of drinking water. However, complete blood count was assessed after an overnight fast. Second, we lacked complete data on job type (13.3% missing) and personal income (4.8% missing), but adjustment for socio-economic position little changed the estimates, suggesting these were not important confounders, so we used dummy categories rather than multiple imputation. Third, the environmental contaminants were not detected directly from the surrounding environment or from blood samples. Instead, we calculated the cumulated contaminants hazards from the questionnaires by multiplying the exposure level and time. So, environmental contaminants may play a role we cannot detect. Fourth,

clearly not all aspects of lifestyle were covered. (26) We did however include a large number of lifestyle factors that are considered to be important to health, such as dietary intake, smoking, exercise and sleep. However, residual confounding due to unmeasured lifestyle is possible. Fifth, there are inherent inaccuracies in nutrient quantitation from food frequency questionnaires because responses may depend on memory, educational level and patience. (24) Social desirability bias may occur if "healthy" food, such as fruits and vegetables, are overstated and "unhealthy" food such as fast food are underreported. However, despite these potential issues, calculated nutrients were generally comparable with those reported using the same instrument in a survey from neighbouring Hong Kong in 1995 (50). We also found the expected associations of vitamin A intake with Hct. Sixth, some of the effect sizes are relatively small and the likelihood of false positive and negative findings cannot be avoided, although we used an FDR of 5% which means that the expected proportion of false positives is less than 5%. Nevertheless the EWAS study has proposed hypotheses which should be tested in future studies, and might shed light on the etiology of CVD, with corresponding implications for prevention and treatment. Small changes at the individual level across the population may result in substantial impact for population health.

There are also some strengths in our study. First, we used a method commonly used in genetic studies, GWAS, but applied it to the "exposome" of environmental influences. Systematic use of all environmental factors limits false positive findings induced by selective reporting of significant results. (51, 52) Second, we included lifestyle factors. (21) We used cross-sectional data systematically to make best use of data on a large set of environmental contaminants, nutritional and major lifestyle factors. Third, data-driven research requires validation of positive findings. We used internal cross-validation with 5 iterations in the present sample. We also found some

associations expected from experimental evidence, which gives greater credence to the other findings.

Such agnostic analyses may identify missing links in important processes, here a potential explanation for harmful effects of vitamin A on mortality by causing cardiovascular disease from raising Hct. From a practical perspective if our findings are confirmed, this study also suggests that vitamin A might be best avoided by people taking anti-coagulants and could be a modifiable exposure positively related to CVD, the role of which in food fortification in well-nourished populations should perhaps be subject to investigation using an unbiased observational method such as Mendelian randomization that avoids potentially harmful interventions.

In conclusion, despite the difficulty of ascertaining causality, our findings that Hct was associated with both expected (vitamin A intake and physical activity) and previously undocumented variables (serum magnesium and serum calcium) suggest EWAS could be useful to generate novel and relevant hypotheses.

Acknowledgements and funding

This work was supported by the Hong Kong Health and Health Services Research Fund (grant 06070981), Health, Welfare and Food Bureau, Government of Hong Kong SAR, People's Republic of China. The Guangzhou Biobank Cohort Study was funded by the University of Hong Kong Foundation for Development and Research (Hong Kong, China), the University of Hong Kong University Research Committee—Strategic Research Theme of Public Health (Hong Kong, China), the Guangzhou Public Health Bureau (Guangzhou, China), the Guangzhou Science and Technology Bureau (Guangzhou, China), the University of Birmingham (Birmingham, United Kingdom), Natural Science Foundation of Guangdong (Guangdong, China)

(grant 9451062001003477), and the Key technology collaboration project funded by the Guangzhou Health Bureau (Grant number: 2012J5100041).

The Guangzhou Biobank Cohort Study investigators include: *Guangzhou No. 12 Hospital*—Dr. W. S. Zhang, Dr. M. Cao, Dr. T. Zhu, Dr. B. Liu, Prof. C. Q. Jiang (Co-Principal Investigator (PI)); *University of Hong Kong*—Dr. C. M. Schooling, Prof. S. M. McGhee, Prof. R. Fielding, Prof. G. M. Leung, Prof. T. H. Lam (Co-PI); *University of Birmingham*—Dr. G. N. Thomas, Dr. P. Adab, Prof. K. K. Cheng (Co-PI).

The funders had no role in the study design, data collection and analysis, the decision to publish, or preparation of the manuscript.

References

1. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. The Lancet. 2012;380(9859):2095-128.

2. Schwartz GG, Olsson AG, Abt M, Ballantyne CM, Barter PJ, Brumm J, et al. Effects of dalcetrapib in patients with a recent acute coronary syndrome. The New England journal of medicine. 2012;367(22):2089-99.

3. Merck. Merck announces HPS2-THRIVE study of Tredaptive™ (Extended-Release Niacin/Laropiprant) did not achieve primary endpoint). 2012.

4. Yousuf O, Mohanty BD, Martin SS, Joshi PH, Blaha MJ, Nasir K, et al. High-sensitivity C-reactive protein and cardiovascular disease: a resolute belief or an elusive link? Journal of the American College of Cardiology. 2013;62(5):397-408.

5. Boussageon R, Bejan-Angoulvant T, Saadatian-Elahi M, Lafont S, Bergeonneau C, Kassaï B, et al. Effect of intensive glucose lowering treatment on all cause mortality, cardiovascular death, and microvascular events in type 2 diabetes: meta-analysis of randomised controlled trials. BMJ. 2011;343.

6. Kanagasabapathy P, Chowdary P, Gatt A. Alternatives to warfarin--the next generation of anticoagulants. Cardiovascular therapeutics. 2011;29(6):e80-8.

7. Burch GE, Depasquale NP. The hematocrit in patients with myocardial infarction. JAMA : the journal of the American Medical Association. 1962;180:62-3.

8. Lee AJ, Mowbray PI, Lowe GD, Rumley A, Fowkes FG, Allan PL. Blood viscosity and elevated carotid intima-media thickness in men and women: the Edinburgh Artery Study. Circulation. 1998;97(15):1467-73.

9. Wannamethee G, Shaper AG, Whincup PH. Ischaemic heart disease: association with haematocrit in the British Regional Heart Study. Journal of epidemiology and community health. 1994;48(2):112-8.

10. Kunnas T, Solakivi T, Huuskonen K, Kalela A, Renko J, Nikkari ST. Hematocrit and the risk of coronary heart disease mortality in the TAMRISK study, a 28-year follow-up. Preventive medicine. 2009;49(1):45-7.

11. Gagnon DR, Zhang TJ, Brand FN, Kannel WB. Hematocrit and the risk of cardiovascular disease--the Framingham study: a 34-year follow-up. Am Heart J. 1994;127(3):674-82.

12. Toss F, Nordstrom A, Nordstrom P. Association between hematocrit in late adolescence and subsequent myocardial infarction in Swedish men. International journal of cardiology. 2013;168(4):3588-93.

13. Ganesh SK, Zakai NA, van Rooij FJ, Soranzo N, Smith AV, Nalls MA, et al. Multiple loci influence erythrocyte phenotypes in the CHARGE Consortium. Nature genetics.

2009;41(11):1191-8.
14. Hashemieh M, Javadzadeh M, Shirkavand A, Sheibani K. Lipid profile in minor thalassemic patients: a historical cohort study. Bangladesh Medical Research Council bulletin. 2011;37(1):24-7.
15. Cocco P, Todde P, Fornera S, Manca MB, Manca P, Sias AR. Mortality in a cohort of men expressing the glucose-6-phosphate dehydrogenase deficiency. Blood. 1998;91(2):706-9.
16. Long WK, Wilson SW, Frenkel EP. Associations between red cell glucose-6-phosphate dehydrogenase variants and vascular diseases. American journal of human genetics. 1967;19(1):35-53.
17. Meloni L, Manca MR, Loddo I, Cioglia G, Cocco P, Schwartz A, et al. Glucose-6-phosphate dehydrogenase deficiency protects against coronary heart disease. Journal of inherited metabolic disease. 2008;31(3):412-7.
18. Coyne DW. The health-related quality of life was not improved by targeting higher hemoglobin in the Normal Hematocrit Trial. Kidney international. 2012;82(2):235-41.
19. Marchioli R, Finazzi G, Specchia G, Cacciola R, Cavazzina R, Cilloni D, et al. Cardiovascular events and intensity of treatment in polycythemia vera. The New England journal of medicine. 2013;368(1):22-33.
20. Patel CJ, Bhattacharya J, Butte AJ. An Environment-Wide Association Study (EWAS) on type 2 diabetes mellitus. PloS one. 2010;5(5):e10746.
21. Jiang C, Thomas GN, Lam TH, Schooling CM, Zhang W, Lao X, et al. Cohort Profile: The Guangzhou Biobank Cohort Study, a Guangzhou–Hong Kong–Birmingham collaboration. International Journal of Epidemiology. 2006;35(4):844-52.
22. Paul AA SD. McCance & Widdowson's: The Composition of Foods. 4th edn ed. London, UK1978.
23. Leung S, Woo J, Ho S, Lam T, Janus E. Hong Kong adult dietary survey, 1995. Australian Journal of Nutrition and Dietetics. 1998.
24. Woo J, Kwok T, Leung J, Tang N. Dietary intake, blood pressure and osteoporosis. J Hum Hypertens. 2008;23(7):451-5.
25. Yang Y WG, Pan G (eds). Institute of Nutrition & Food Safety, China CDC, In: China Food Composition 2002: University Medical Press: Peking, China; 2002.
26. Lind PM, Riserus U, Salihovic S, Bavel B, Lind L. An environmental wide association study (EWAS) approach to the metabolic syndrome. Environment international. 2013;55:1-8.
27. Smith JC, Makdani D, Hegar A, Rao D, Douglass LW. Vitamin A and zinc supplementation of preschool children. Journal of the American College of Nutrition. 1999;18(3):213-22.

28. Tanumihardjo SA, Permaesih D, Muhilal. Vitamin A status and hemoglobin concentrations are improved in Indonesian children with vitamin A and deworming interventions. European journal of clinical nutrition. 2004;58(9):1223-30.
29. Chen K, Zhang X, Li TY, Chen L, Wei XP, Qu P, et al. Effect of vitamin A, vitamin A plus iron and multiple micronutrient-fortified seasoning powder on infectious morbidity of preschool children. Nutrition (Burbank, Los Angeles County, Calif). 2011;27(4):428-34.
30. Mejia LA, Chew F. Hematological effect of supplementing anemic children with vitamin A alone and in combination with iron. The American journal of clinical nutrition. 1988;48(3):595-600.
31. Howard BJ, Fraser SF, Sethi P, Cerin E, Hamilton MT, Owen N, et al. Impact on hemostatic parameters of interrupting sitting with intermittent activity. Medicine and science in sports and exercise. 2013;45(7):1285-91.
32. Romain AJ, Brun JF, Varlet-Marie E, Raynaud de Mauverger E. Effects of exercise training on blood rheology: a meta-analysis. Clinical hemorheology and microcirculation. 2011;49(1-4):199-205.
33. Varlet-Marie E, Brun JF, Fedou C, Raynaud de Mauverger E. Blood rheology and body composition as determinants of exercise performance in male soccer players. Clinical hemorheology and microcirculation. 2011;49(1-4):225-30.
34. Mairbaurl H. Red blood cells in sports: effects of exercise and training on oxygen supply by red blood cells. Frontiers in physiology. 2013;4:332.
35. Villamor E, Fawzi WW. Effects of vitamin a supplementation on immune responses and correlation with clinical outcomes. Clinical microbiology reviews. 2005;18(3):446-64.
36. Semba RD, Bloem MW. The anemia of vitamin A deficiency: epidemiology and pathogenesis. European journal of clinical nutrition. 2002;56(4):271-81.
37. Bhasin S, Storer TW, Berman N, Callegari C, Clevenger B, Phillips J, et al. The Effects of Supraphysiologic Doses of Testosterone on Muscle Size and Strength in Normal Men. New England Journal of Medicine. 1996;335(1):1-7.
38. Wittert GA, Chapman IM, Haren MT, Mackintosh S, Coates P, Morley JE. Oral Testosterone Supplementation Increases Muscle and Decreases Fat Mass in Healthy Elderly Males With Low–Normal Gonadal Status. The Journals of Gerontology Series A: Biological Sciences and Medical Sciences. 2003;58(7):M618-M25.
39. Bjelakovic G, Nikolova D, Gluud LL, Simonetti RG, Gluud C. Mortality in randomized trials of antioxidant supplements for primary and secondary prevention: systematic review and meta-analysis. JAMA : the journal of the American Medical Association. 2007;297(8):842-57.
40. Myung S-K, Ju W, Cho B, Oh S-W, Park SM, Koo B-K, et al. Efficacy of vitamin and

antioxidant supplements in prevention of cardiovascular disease: systematic review and meta-analysis of randomised controlled trials2013 2013-01-18 12:14:28.
41. El-Sayed MS, El-Sayed Ali Z, Ahmadizad S. Exercise and training effects on blood haemostasis in health and disease: an update. Sports medicine (Auckland, NZ). 2004;34(3):181-200.
42. Weber JM, Forsythe SR, Christianson CA, Frisch BJ, Gigliotti BJ, Jordan CT, et al. Parathyroid hormone stimulates expression of the Notch ligand Jagged1 in osteoblastic cells. Bone. 2006;39(3):485-93.
43. Shiozawa Y, Jung Y, Ziegler AM, Pedersen EA, Wang J, Wang Z, et al. Erythropoietin couples hematopoiesis with bone formation. PloS one. 2010;5(5):e10853.
44. Teo KK, Yusuf S, Collins R, Held PH, Peto R. Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomised trials. Bmj. 1991;303(6816):1499-503.
45. Li J, Zhang Q, Zhang M, Egger M. Intravenous magnesium for acute myocardial infarction. The Cochrane database of systematic reviews. 2007(2):Cd002755.
46. Chan-Yeung M, Ferreira P, Frohlich J, Schulzer M, Tan F. The effects of age, smoking, and alcohol on routine laboratory tests. American journal of clinical pathology. 1981;75(3):320-6.
47. Shaper AG, Pocock SJ, Ashby D, Walker M, Whitehead TP. Biochemical and haematological response to alcohol intake. Annals of clinical biochemistry. 1985;22 (Pt 1):50-61.
48. Thun MJ, Hannan LM, Adams-Campbell LL, Boffetta P, Buring JE, Feskanich D, et al. Lung cancer occurrence in never-smokers: an analysis of 13 cohorts and 22 cancer registry studies. PLoS medicine. 2008;5(9):e185.
49. Li Q, Hsia J, Yang G. Prevalence of Smoking in China in 2010. New England Journal of Medicine. 2011;364(25):2469-70.
50. Woo J. Nutrition and health issues in the general Hong Kong population. Hong Kong medical journal = Xianggang yi xue za zhi / Hong Kong Academy of Medicine. 1998;4(4):383-8.
51. Blair A, Saracci R, Vineis P, Cocco P, Forastiere F, Grandjean P, et al. Epidemiology, public health, and the rhetoric of false positives. Environmental health perspectives. 2009;117(12):1809-13.
52. Young S. Acknowledge and fix the multiple testing problem. International Journal of Epidemiology. 2010;39(3):934.

Appendix Table A: Means and medians for the variables from the Guangzhou Biobank Cohort Study.

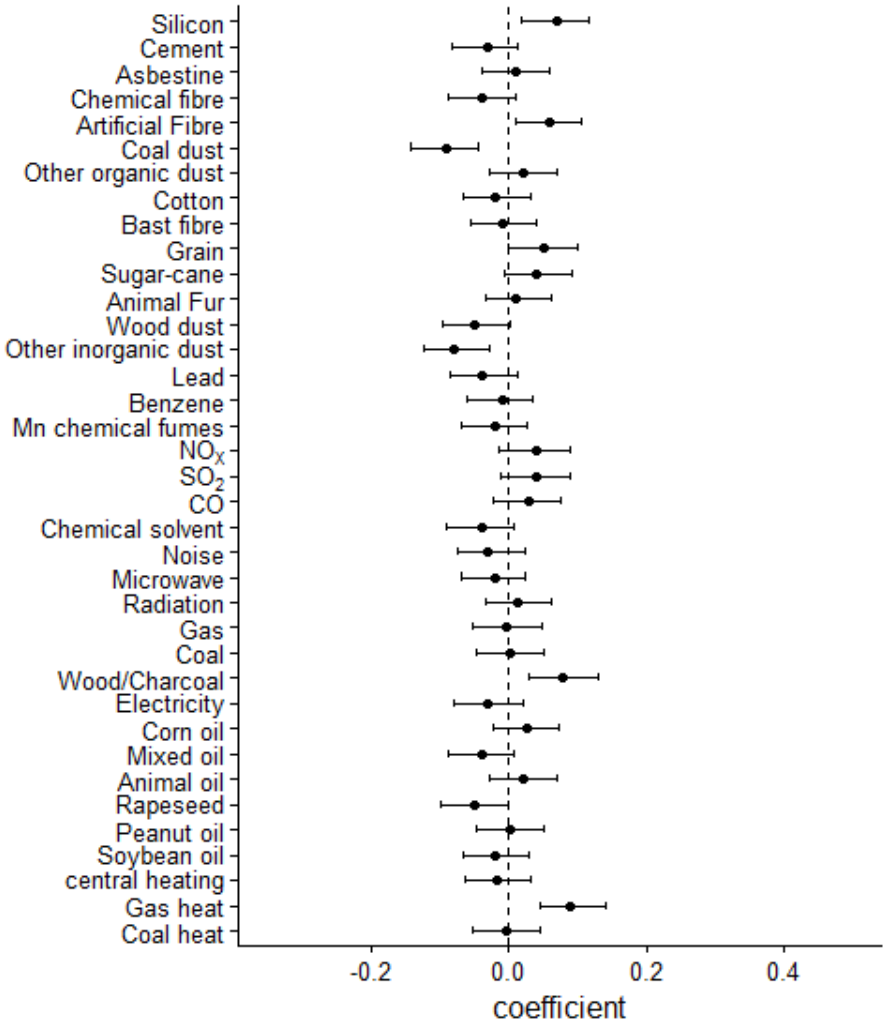
Variable	N	Mean (SD)	Median (25^{th} and 75^{th}
Dietary factors			
Energy (Kcal/day)	20404	1833.4(534.8)	1788.3 (1458.7,2151.4)
Protein (% of total energy intake)	20404	15.9 (3.0)	15.8 (14.0, 17.7)
Carbohydrates (% of total energy intake)	20404	57.0 (9.0)	57.2 (51.2, 63.0)
Fat (% of total energy intake)	20404	28.6 (8.7)	28.1 (22.6, 34.1)
Fibrin (g/day)	20404	13.2 (6.7)	11.9 (8.8, 16.2)
Cholesterol (g/day)	20404	149.2 (98.3)	135.5 (94.3, 188.4)
Saturated fatty acid (SFA) (g/day)	20404	10.1 (5.2)	9.3 (6.4, 12.8)
Monounsaturated fatty acid (MUFA) (g/day)	20404	17.3 (9.0)	16.2 (11.0, 22.2)
Polyunsaturated fatty acid (PUFA) (g/day)	20404	13.0 (8.4)	12.3 (7.2, 17.1)
Vitamin A (VitA) (ug/day)	20404	621.7 (502.7)	491.3 (324.5, 755.7)
Vitamin B1 (VitB1) (mg/day)	20404	1.02 (0.41)	0.96 (0.73, 1.23)
Vitamin B2 (VitB2) (mg/day)	20404	0.96 (0.40)	0.90 (0.71, 1.14)
Niacin (mg/day)	20404	23.5 (7.5)	22.7 (18.4, 27.6)
Vitamin C (VitC) (mg/day)	20404	97.9 (65.5)	82.8 (57.8, 120.7)
Vitamin E (VitE) (mg/day)	20404	18.5 (9.76)	17.3 (11.8, 23.5)
Calcium (Ca) (mg/day)	20404	504.9 (240.4)	459.2(333.9,623.5)
Phosphorous (P) (mg/day)	20404	1083.5 (352.6)	1041.2 (840.1, 1277.4)
Iron (Fe) (mg/day)	20404	22.4 (7.86)	21.4 (16.9, 26.6)
Magnesium (Mg) (mg/day)	20404	305.3 (133.3)	275.8 (215.6, 361.4)
Sodium (Na) (mg/day)	20404	937.5 (653.0)	840.7 (594.9, 1168.1)
Potassium (K) (mg/day)	20404	1669.2 (648.9)	1041.2 (840.1, 1277.4)
Zinc (Zn) (mg/day)	20404	11.7 (3.6)	11.2 (9.2, 13.6)
Copper (Cu) (mg/day)	20404	3.79 (4.34)	2.06 (1.33, 4.78)
Selenium (Se) (ug/day)	20404	50.3 (21.5)	47.0 (35.9, 61.0)
Manganese (Mn) (mg/day)	20404	5.63 (2.39)	5.15 (4.00,6.73)
Environmental contaminants			
Dust			
Silicon	19024	1.66 (9.89)	
Cement	19024	1.85 (10.1)	
Asbestine	19024	0.36 (4.27)	
Chemical fibre	19024	0.17 (2.91)	
Artificial fibre	19024	0.35 (4.37)	
Coal dust	19024	0.97 (7.54)	
Other organic dust	19024	6.3 (18.4)	
Cotton	19024	1.9 (10.3)	
Bast fibre	19024	0.16 (3.01)	
Grain	19024	1.26 (7.94)	
Sugar-cane	19024	0.33 (4.51)	
Animal fur	19024	0.04 (1.33)	
Wood	19024	0.4 (4.77)	
Other inorganic dust	19024	1.51 (9.01)	
Chemical fumes			
Lead	20415	0.29 (4.15)	
Benzene/Toluene/Dimethylbenzene	20415	0.79 (6.71)	

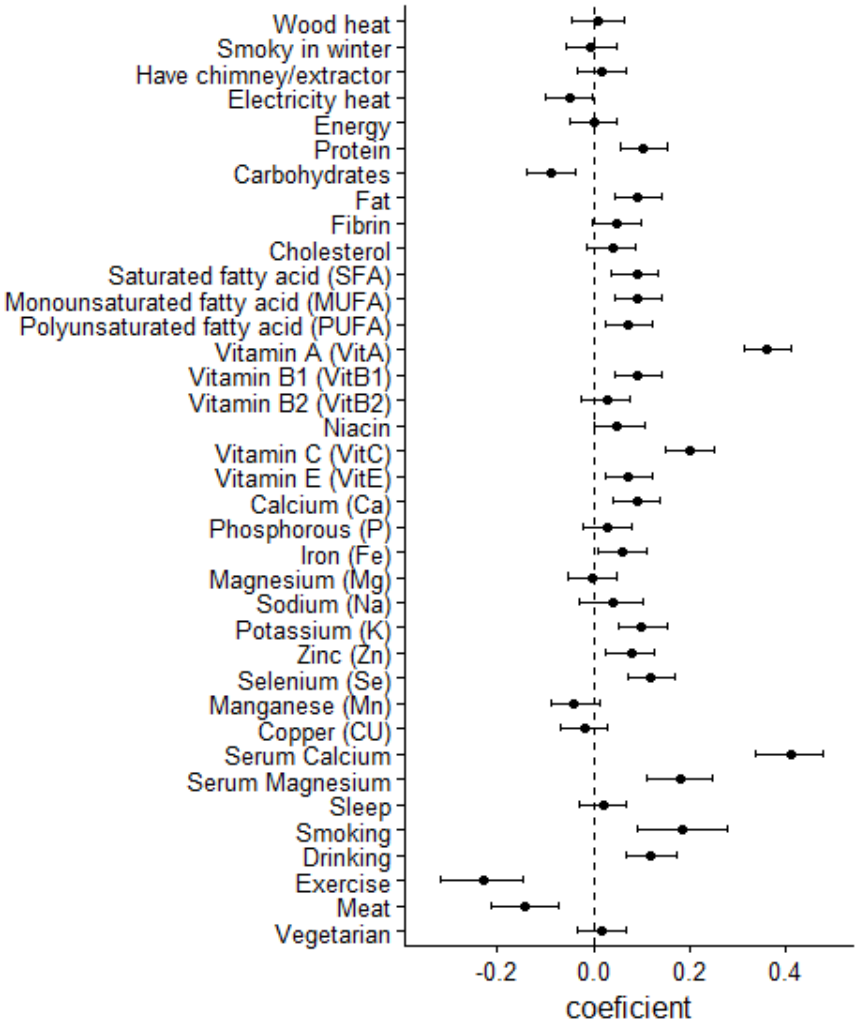
Manganese (Mn)	20415	0.23 (3.63)	
Sulfur dioxide (SO_2)	20415	0.37 (4.91)	
Nitrogen oxide species (NOx)	20415	0.12 (2.52)	
Carbon monoxide (CO)	20415	0.19 (3.46)	
Chemical solvent	20415	2.18 (11.1)	
Physical exposures			
Noise	20415	9.72 (25.0)	
Microwave	20415	0.08 (2.47)	
radiation	20415	0.36 (4.82)	
Fuel for cooking			
Gas	20403	22.74 (15.4)	
Coal	20409	7.16 (12.4)	
Wood/charcoal	20410	3.56 (10.2)	
electricity	20415	0.22 (2.5)	
Oil for cooking			
Corn oil	20400	4.76 (11.6)	
Mixed oil	20414	1.09 (6.15)	
Animal oil	20411	0.45 (3.80)	
Rapeseed	20412	0.77 (4.87)	
Peanut	20331	42.3 (21.2)	
Soy bean	20415	0.36 (3.43)	
Fuel for heat			
Central heating	20413	0.07 (1.38)	
Gas	20415	0.01 (0.62)	
Coal	20415	0.07 (1.48)	
Wood/charcoal	20415	0.05 (1.26)	
Electricity	20412	4.14 (10.8)	
Have chimney/extractor	20179	6.75 (13.3)	
Smoky in winter	20162	3.84 (10.7)	
Blood metals			
Serum Calcium (mmol/L)	10408	2.16	2.14 (2.05, 2.24)
Serum Magnesium (mmol/L)	10405	0.94	0.91 (0.81, 1.04)
Lifestyle related factors			
Hours slept per night (sleep)	20027	6.92 (1.3)	7.0 (6.0, 8.0)
Smoking status	19024		
Never		80.4%	
Ex-smoker		9.7%	
Current		9.9%	
Drinking status	20443		
Never		82.1%	
<1/week		2.5%	
1-4/week		7.1%	
5+/week		1.6%	
Ex-drinker		2.4%	
Unknown		4.3%	
Physical activity[a]	19024		
Inactive		8.1%	
Minimally active		47.2%	
HEPA[b]		44.7%	

Meat eating habits	9193	
Never		0.8%
Yearly		14.2%
Monthly		26.4%
Weekly		38.2%
Daily		10.3%
Vegetarian	20397	0.8%

[a]Assessed with the International Physical Activity Questionnaire

[b]Health enhancing physical activity—that is, vigorous activity at least 3 days a week that corresponds to a minimum of 1500 metabolic equivalent (MET) minutes per week, or activity 7 days of the week that corresponds to at least 3000 MET minutes per week.


Silicon
Cement
Asbestine
Chemical fibre
Artificial Fibre
Coal dust
Other organic dust
Cotton
Bast fibre
Grain
Sugar-cane
Animal Fur
Wood dust
Other inorganic dust
Lead
Benzene
Mn chemical fumes
NO_X
SO_2
CO
Chemical solvent
Noise
Microwave
Radiation
Gas
Coal
Wood/Charcoal
Electricity
Corn oil
Mixed oil
Animal oil
Rapeseed
Peanut oil
Soybean oil
central heating
Gas heat
Coal heat
-0.2
0.0
0.2
0.4
coefficient


Wood heat
Smoky in winter
Have chimney/extractor
Electricity heat
Energy
Protein
Carbohydrates
Fat
Fibrin
Cholesterol
Saturated fatty acid (SFA)
Monounsaturated fatty acid (MUFA)
Polyunsaturated fatty acid (PUFA)
Vitamin A (VitA)
Vitamin B1 (VitB1)
Vitamin B2 (VitB2)
Niacin
Vitamin C (VitC)
Vitamin E (VitE)
Calcium (Ca)
Phosphorous (P)
Iron (Fe)
Magnesium (Mg)
Sodium (Na)
Potassium (K)
Zinc (Zn)
Selenium (Se)
Manganese (Mn)
Copper (CU)
Serum Calcium
Serum Magnesium
Sleep
Smoking
Drinking
Exercise
Meat
Vegetarian
-0.2
0.0
0.2
0.4
coeficient

Appendix Fig A. Forest plot showing the coefficient and 95%CI for the associations between z-scores of 74 nutrient elements, environment contaminants or lifestyle factors with Hct in the Guangzhou Biobank Cohort, Guangzhou Study, China, 2003-2006.